KOUTILYA PANDE

New Brunswick, NJ |+1 (929) 3785340 | koutilyapande26@gmail.com | LinkedIn

Data Scientist with 2 years of experience building end-to-end machine learning models from 0 to 1. Skilled in Python, SQL, statistical analysis, model training and data visualization to streamline insights from large datasets, parallely exploring the realm of AI with RAG and Agentic Wokflows. Looking to collaborate with Data Scientists and AI Research teams to develop unique, practical, and scalable solutions to data-driven challenges.

SKILLS

Languages: C, C++,R, Python (Pandas, NumPy, Matplotlib, TensorFlow, Scikit Learn, Keras), SQL
Functional: ETL, Statistical Analysis, Data Visualization, ML Algorithms, NLP, Conversational AI, RAG, AI Agents
Database: MS SQL, My SQL, MongoDB Atlas, NoSQL, Vector Database (Pinecone, ChromaDB, Bedrock)
Tools and Platforms: MS Excel, Tableau Desktop, PowerBI, Jupyter, Azure AI, GitHub, Huggingface, LLMs

EXPERIENCE

Data Scientist ***Feb'24 – Present***
Keelworks Foundation Oak Harbor,USA

- Led data collection and data preprocessing team, ensuring the data quality and integrity for models
- Implemented **sentiment analysis** using **BERT** and **NER** using **spaCy** with fellow machine learning engineers for improving customer feedback categorization and enhanced information extraction efficiency by **40**%
- Optimized homepage by conducting **A/B testing** and **statistical analysis**, resulting in a **28%** increase in **CTR**
- Tracked and analyzed core business KPIs using **PowerBI** and **Excel**,highlighting insights on user behavior

Data Scientist ***Dec'21 – Jul'22***
Korangle India

- Automated **data sampling** using **Python**, reducing processing time by 3**0%** while conducting rigorous **statistical analyses** on 50,000+ entries for data validation, ensuring **data integrity** for development and marketing strategies
- Developed a customer segmentation model using **K-Means clustering** to find new set of customers
- Analyzed competitive market data and customer queries using **NLP**, increasing market penetration by **17%**
- Assisted with customer journey mapping, tested **Random Forest** customer satisfaction model from scratch, synthesizing data from customer surveys and reviews, resulting in **12%** fewer customer complaints

Data Science Analyst ***May'21 – July'21***
Udemy India

- Managed and optimized **ETL** pipeline by leveraging **SQL** and **Python**, reducing data preparation time by **18%**
- Designed complex queries, stored procedures, triggers and cursors in **SQL Server** to extract structured data for business metrics generation using matplotlib, eliminating **8hr/week** of manual work
- Devised feature engineering and cross-validated **classification models** to predict loan approval outcomes
- Successfully addressed **class imbalance** by conducting **model evaluation**(confusion matrix analysis, AUC-ROC), and fine-tuned to increase model accuracy by **16%**

PROJECTS

AI Agentic Workflows ***Jan'25-Present***

- Essay Grading Agent | GitHub | *(LangGraph, OpenAI, Langsmith)*
- Financial Agent with Web Search News Tool | GitHub | *(phidata Tavily, FASTAPI)*

PDF ChatGenie | GitHub **|** *(RAG, Streamlit, LangGraph, LangChain)* ***June'24 – July'24***

- Developed a multi-PDF **conversational chatbot** application for providing contextual aware answers through RAG and using Pinecone for vector storage and retrival with chat buffer memory

Naruto Series Analyzer - NLP | Github **|** *(Transformers, HuggingFace, LORA, Gradio, Llama)* ***Feb'24-Apr'24***

- Built a Gradio web application using SpaCy for NER and BERT-based **zero-shot classifiers** to analyze character interactions and themes from scraped datasets using Scrapy
- Embedded custom jujutusu classifiers using **DistilBERT** and a Naruto chatbot using Transformers and Huggingface, fine-tuning **Llama 3.2** using LORA

Dodging The Mishap - Web Application | GitHub **|** *(Time-Series, LSTM, Django, YFinance)* ***Jan'23-Apr'23***

- Designed a stock price prediction Django web application powered by an **LSTM** machine learning model, supporting **time series analysis** on real time financial data from YFinance

MoodSync: Speech-Driven Spotify Song Recommendations | GitHub **|** *(Python, Tensorflow, sklearn)* ***Jan'23-Apr'23***

- Authored a **Recommendation System** deploying Multi-Layer Perceptron Classifier from scrath to recognize emotions through speech Librosa and recommended Spotify playlists based on identified emotion

EDUCATION

RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY ***Sept'22 – Dec'23***
Master of Information Technology and Analytics, Major :- ***Data Science*** ***3.7/4.0***
SHRI GOVINDRAM SEKSARIA INSTITUTE OF TECHNOLOGY AND SCIENCE, INDIA ***Aug'18 – May'22***
Bachelor of Technology (B. Tech) : Electronics and Telecommunication Engineering